Marker-assisted Selection in Backcross Breeding

S.J. Openshaw

Pioneer Hi-Bred Intl. Inc., P.O. Box 1004, Johnston, IA 50131

S.G. Jarboe[1]

CIMMYT, Lisboa 27, Apdo. Postal 6-641, 06600 Mexico, D.F., Mexico

W.D. Beavis

Pioneer Hi-Bred Intl. Inc., P.O. Box 1004, Johnston, IA 50131

Abstract. **The backcross breeding procedure has been used widely to transfer simply inherited traits into elite genotypes. Genetic markers can increase the effectiveness of backcrossing by 1) increasing the probability of obtaining a suitable conversion, and 2) decreasing the time required to achieve an acceptable recovery. Simulation and field results indicated that, for a genome consisting of ten 200-cM chromosomes, basing selection on 40 or 80 markers in 50 BC individuals that carry the allele being transferred can reduce the number of backcross generations needed from about seven to three.**

The backcross breeding procedure has been used widely to transfer simply inherited traits into elite genotypes. Usually, the trait being transferred is controlled by a single gene, but highly heritable traits that are more complexly inherited have also been transferred successfully by backcrossing; for example, maturity in maize (Rinke and Sentz, 1961; Shaver, 1976). Today, backcrossing is being used to transfer genes introduced by such techniques as transformation or mutation into appropriate germplasm.

Several plant breeding textbooks give good descriptions of the backcross procedure (Allard, 1960; Fehr, 1987). A donor parent (DP) carrying a trait of interest is crossed to the recurrent parent (RP), an elite line that is lacking the trait. The F_1 is crossed back to the RP to produce the BC_1 generation. In the BC_1 and subsequent backcross generations, selected individuals carrying the gene being transferred are backcrossed to the RP. The expected proportion of DP genome is reduced by half with each generation of backcrossing. Ignoring effects of linkage to the selected DP allele being transferred, the percentage recurrent parent (%RP) genome expected in each backcross generation is calculated as:

$$\%RP = 100\,[1 - (0.5)^{n+1}]$$

where n is the number of backcrosses.

Backcrossing of selected plants to the RP can be repeated each cycle until a line is obtained that is essentially a version of the RP that includes the introgressed allele. After six backcrosses, the expected recovery is >99% (Table 1).

Until recently, discussions of the recovery of the RP genome during backcrossing have emphasized the expected values for %RP shown in Table 1, and have largely ignored the genetic variation for %RP that exists around the expected mean. With the development of genetic markers capable of providing good genome coverage, there has been interest in taking advantage of that variation to increase the efficiency of backcrossing.

Selection for RP marker alleles can increase greatly the effectiveness of backcross programs by allowing the breeder to 1) select backcross plants that have a higher proportion of RP genome, and 2) select backcross individuals that are better conversions near a mapped donor allele being transferred (i.e., select for less linkage drag). Expressed in practical terms, using genetic markers to assist backcrossing can 1) increase the probability of obtaining a suitable conversion, and 2) decrease the time required to achieve an acceptable recovery.

Issues to consider when planning a marker-assisted backcross program include 1) the time advantage of using markers to assist backcrossing, 2) the number of markers needed, and 3) the number of genotypes to evaluate. In this report, we use results from previous literature, computer simulation, and empirical studies to provide some guidelines.

Table 1. *Expected recovery of recurrent parent (RP) genome during backcrossing, assuming no linkage to the gene being transferred.*

Generation	%RP
F_1	50.0000
BC_1	75.0000
BC_2	87.5000
BC_3	93.7500
BC_4	96.8750
BC_5	98.4375
BC_6	99.2188
BC_7	99.6094

[1]Formerly with Purdue University, West Lafayette, Ind.

Materials and methods

The maize genome was the model for the simulation. The simulated genome contained ten 200-cM chromosomes. Simulation of crossing over was based on a Poisson distribution with a mean of 2.0 ($\lambda = 2$) (Hanson, 1959), which, on average, generated one cross over for every 100-cM length. The simulations reported here assume no interference. Codominant genetic markers were evenly distributed in the genome and sites of the donor gene were randomly assigned to genome locations.

Simulations were conducted with the following parameters:

Number of progeny: 100 or 500.
Backcross generations: BC_1, BC_2, and BC_3.
Number of markers: 20, 40, 80, or 100.
Number selected to form the next BC generation: 1 or 5.

Selection was based on 1) presence of the donor allele and 2) high %RP). %RP was calculated as the average of the (one or five) selected individuals. Values presented are the mean of 50 simulations.

Results

In the computer simulation study, all methods modeled greatly increased the speed of recovering the RP genome compared to the expected recovery with no marker-assisted selection (compare Tables 1 and 2). At least 80 markers were required to recover 99% of the RP genome in just three BC generations (Table 2). Use of at least 80 markers and 500 progeny allowed recovery of 98% RP in just two BC generations. Response to selection was diminished only slightly by spreading the effort over five selections. Using markers, the number of backcross generations needed to convert an inbred is reduced from about seven to three.

By the BC_2 generation, there appears to be no practical advantage to using 500 vs. 100 individuals. If the presence of the donor trait in the backcross individuals can be ascertained before markers are genotyped, then only half the number of individuals indicated in the tables will need to be analyzed.

When a small number f markers are used, they quickly became non-informative; i.e., selection causes the marker loci to became fixed for the RP type before the rest of the genome is fully converted (Table 3; Hospital et al., 1992). This situation was most prominent in the larger populations, where a higher selection intensity placed more selection pressure upon the marker loci. Accordingly, it is of interest to consider how closely the estimation of %RP based on markers reflects the actual genome composition. The combination of estimation of %RP based on fewer markers and subsequent selection tends to bias the estimates upward (compare Tables 2 and 3).

The results from the simulation compare well with real field data. In a typical example, 50 BC_1 plants carrying the gene being transferred were genotyped at 83 polymorphic RFLP loci (note that this corresponds to a population size of 100 unselected plants in Tables 2 and 3). The five best BC_1 recoveries had estimated %RP values of 85.9%, 82.7%, 82.0%, 81.4%, and 81.2%. After evaluating 10 BC_2 plants from each selected BC_1, the best BC_2 recovery had an estimated %RP of 94.6%.

Discussion

The simulations (Table 2; Hospital et al., 1992) and our experience indicate that four markers per 200-cM chromosome is adequate to greatly increase the effectiveness of selection in the BC_1. However, using only four markers per 200 cM will likely make it very difficult to map the location of the gene of interest. Adequate summarization of the data is an important

Table 2. *Percent recurrent parent genome during marker-assisted backcrossing.*

	100 Progeny				500 Progeny			
	No. markers				No. markers			
Generation	20	40	80	100	20	40	80	100
				One selected				
BC_1	84.5	84.5	84.2	88.0	89.9	90.7	90.2	90.5
BC_2	95.0	95.2	95.8	97.2	96.5	97.7	98.5	98.6
BC_3	97.4	97.6	98.9	99.2	97.7	98.3	99.4	99.5
				Five selected				
BC_1	82.9	85.1	84.9	84.7	87.7	88.1	88.9	88.9
BC_2	93.7	95.0	95.8	95.7	95.5	96.8	97.8	97.9
BC_3	97.1	98.3	98.8	98.9	97.3	98.5	99.3	99.3

Table 3. *Estimates of percent recurrent parent genome, based on marker loci.*

	100 Progeny				500 Progeny			
	No. markers				No. markers			
Generation	20	40	80	100	20	40	80	100
				One selected				
BC_2	98.7	97.8	95.6	97.2	100.0	99.1	98.6	98.0
BC_3	100.0	99.8	99.3	99.5	100.0	100.0	99.9	98.2
				Five selected				
BC_2	96.4	96.5	96.2	95.8	100.0	98.5	98.3	98.2
BC_3	99.9	99.8	99.3	99.1	100.0	100.0	99.9	99.8

part of a marker-assisted backcross program. Ideally, the markers used can supply data that can be represented as alleles of loci with known map position. Estimation of %RP, mapping the position of the locus of interest, and graphical display of the results (Young and Tanksley, 1989) are all useful in understanding and controlling the specific backcross experiment being conducted.

It appears that, with the use of genetic markers, the portion of the RP genome that is not linked to the allele being transferred can be recovered quickly and with confidence. The recovery of RP will be slower on the chromosome carrying the gene of interest. A considerable amount of linkage drag is expected to accompany selection for the DP allele in a backcross program. For a locus located in the middle of a 200-cM chromosome, the length of the DP chromosome segment accompanying selection is expected to be 126, 63, and 28 cM in the BC_1, BC_2, and BC_3 generations, respectively (Hanson, 1959; Naveira and Barbadilla, 1992). Our observations support the recommendation of Hospital et al. (1992) that preference be given to the selection for recombinants proximal to the allele of interest, but that selection for recovery of the RP elsewhere in the genome also be considered. This two-stage selection can probably be done quite effectively ad hoc by the breeder once the data is adequately summarized; however, Hospital et al. suggest ways to incorporate the two criteria into a selection index such that each component of selection is assured appropriate weighting.

Use of genetic markers can greatly increase the effectiveness of backcrossing, and they should be used in any serious backcrossing program if resources are available to the breeder.

Literature Cited

Allard, R.W. 1960. Principles of plant breeding. Wiley, New York.

Fehr, W.F. 1987. Principles of cultivar development: v.1. Theory and technique. Macmillan, New York.

Hanson, W.D. 1959. Early generation analysis of length of heterozygous chromosome segments around a locus held heterozygous with backcrossing or selfing. Genetics 44:843–847.

Hospital, F., C. Chevalet, and P. Mulsant. 1992. Using markers in gene introgression breeding programs. Genetics 132:1199–1210.

Rinke, E.H and J.C. Sanz. 1961. Moving corn-belt germplasm northward. Ann. Hybrid Corn Industry Conf. 16:53–56.

Shaver, D.L. 1976. Conversions for earliness in maize inbreds. Maize Genet. Coop. Nwsltr. 50:20–23.

Young, N.D. and S.D. Tanksley. 1989. Restriction fragment length polymorphism maps and the concept of graphical genotypes. Theo. Applied Genet. 77: 95–101.